**Proof on FDR control from Benjamini & Hochberg (JRSSB, 1995)**



**The procedure.** Consider testing H1, H2, …, Hm based on the corresponding p-values P1, P2, …, Pm. Let P(1)≤P(2)≤…≤P(m) be the ordered p-values, and denote by H(i) the null hypothesis corresponding to P(i). Define the following Bonferroni-type multiple-testing procedure:

Let k be the largest i for which 

Then reject all H(i), i=1,2,…,k.

**Lemma**. For any 0≤m0≤m independent p-values corresponding to true null hypotheses, and for any values that the m1=m-m0 p-values corresponding to the false null hypotheses can take, the multiple-testing procedure defined by procedure above satisfies the inequality

  (1)

Proof. This lemma can be proved by induction.

When m=1, m0=0 implies Q≡0, then both sides of (1) are 0s;

 m0=1 implies E(Q)=E{I{P1 q\*}]=Pr(P1 q\*)=q\*=RH of (1).

Assume that the lemma is true for all *k≤m*, and will show that it holds for m+1.

If m0=0, obviously Q≡0 and (1) holds.

If m0>0, denote by {Pi’, i=1, 2 ,…, m0}, the p-values corresponding to the true null hypotheses, and the largest of these by P(m0)’. These are U(0,1) independent random variables. For ease of notation assume that the m1 p-values that the false null hypotheses take are ordered p1≤p2≤…≤pm1. Then let j0 be the largest 0≤j≤ m1 satisfying

  (2)

and denote the right hand side of (2) at j0 by .

Conditioning on P(m0)’=p,

 (3)
Because P(m0)’ is the largest one of m0 independent U(0,1) variables, we have

 

In the first part of (3), p≤p”. Then among the p-values for the m0 nulls and the j0 smallest p-values among the false nulls, their maximum is smaller than , but .

Therefore only the first (m0+ j0) hypotheses would be rejected and V= m0 and .

Then

For the second component of (3), consider two scenarios:

1. 
2. 

Under (a), , therefore the true null corresponding to P(m0)’ and the false nulls corresponding to p(j+1),…,pm1 all would be accepted. The rejection or acceptance of the other hypotheses would not depend on the actual values of P(m0)’=p, p(j+1), …, pm1. The same is true for (b) as well, except that j is replaced by j0.

Therefore for the remain (m0+j-1) hypotheses, the procedure is to find out the largest k≤ m0+j-1, such that P(k)≤{k/(m+1)}q\*, where P(k) is the kth smallest p-value from the set

 {Pi’, i=1, 2 ,…, m0-1} U { p1,p2, …, pj}

Or equivalently

 

Because condition on P(m0)’=p, {Pi’/p, i=1, 2 ,…, m0-1} are (m0-1) independently distributed U(0,1) variables,0≤ pi/p≤1 are numbers corresponding to false null hypotheses. Then the false discovery rate Q is the same as the one from controlling the false discovery rate among (m0-1) true nulls and j false nulls with p-values { Pi’/p, i=1, 2 ,…, m0-1; pi/p, i=1,2,…,j} with false discovery bound by 

Applying the induction, we have



Summing the two parts, we got



End of proof.